To: Hessling, Michael[Hessling.Michael@epa.gov];
drupal_oei_support[drupal_oei_support@epa.gov]
Cc: WCSD Webteam[WCSD_Webteam@epa.gov]
From: Spradling, LeAnn
Sent: Wed 2/8/2017 9:21:07 PM
Subject: RE: Please move *.conf files to snapshot

OK, I am beginning to see what is going on here. I'm afraid it is going to be very complicated but I will simplfy as much as possible to save everyone time.

Here is one example for us to walk through, from application_server_redirects.conf.

RedirectMatch 301 ^/sv/?$
http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do

RedirectMatch 301 ^/sv/(.*)
http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do

Testing that one:
WORKS: https://19january2017snapshot.epa.gov/sv/whatever ->
https://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do.
WORKS: https://19january2017snapshot.epa.gov/sv/ ->
https://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do
--> NO WORK: https://19january2017snapshot.epa.gov/sv ->
https://19january2017snapshot.epa.gov/sv_.html

What is happening there on the NO WORK is a result of the very first set of rewrites we put in. We put in a rewrite block which says that if the request does NOT end with a slash and does not end with .something (like .html, .pdf, etc) rewrite to whatever_.html. It looks like this:

RewriteEngine On
RewriteCond %{REQUEST_URI} !^/www3/.*$
RewriteCond %{REQUEST_URI} !/$
RewriteCond %{REQUEST_URI} !^(.*)\.(.*)$
RewriteRule ^(.*)$ $1_.html [R]

I need to do some more investigation here but it would seem what we have run into is a classic conflicting rewrite rule problem. When Apache sees
Snapshot.epa.gov/sv

It picks the best option (I am oversimplifying a lot here so as not to jam up everyone's email). In this case there are several options for it to look at:

Does it have something after the slash ie https://19january2017snapshot.epa.gov/sv/whatever
Does it have an ending slash? If yes then Redirect to
http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do
If no ending slash, then there are two possible rules that come into play:
RedirectMatch 301 ^/sv/?$
http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/sciencevocabulary/search.do
OR
RewriteCond %{REQUEST_URI} !^(.*)\.(.*)$
As we see it picks the last one and redirects to https://19january2017snapshot.epa.gov/sv_.html. Most likely this is a precedence issue - that is the first rule in the config so it is the one that gets obeyed.

There are 8426 active redirects in the *conf files that have been moved over. We can think of that as

8426 potential conflicts. We may have a problem like this with every URL that does NOT end with a / and does NOT end with .something.

I need to cogitate and see if there is a simple way to solve this. Fixing this problem could take a very large amount of time. Any suggestions?

Thanks,
LeAnn Spradling | Director of Consulting – Service Delivery Manager
CAP/RAP EPA CAM | CGI Federal

T: [Ex. 6 - Personal Privacy] |M: [Ex. 6 - Personal Privacy] | spradling.leann@epa.gov

-----Original Message-----
From: Spradling, LeAnn
Sent: Wednesday, February 08, 2017 3:54 PM
To: Hessling, Michael <Hessling.Michael@epa.gov>; drupal_oei_support <drupal_oei_support@epa.gov>
Cc: WCSD Webteam <WCSD_Webteam@epa.gov>
Subject: RE: Please move *.conf files to snapshot

This is interesting. Those redirects are comment out in the *conf file that I brought over

drupal_redirect.conf:#RedirectMatch 301 ^/climatechange/science/?$ ^/climate-change-science
drupal_redirect.conf:#RedirectMatch 301 ^/climatechange/science/(.*) ^/climate-change-science

How is this being done on the actual www.epa.gov servers? The redirect table?

Thanks,
LeAnn Spradling | Director of Consulting – Service Delivery Manager CAP/RAP EPA CAM | CGI Federal

T: [Ex. 6 - Personal Privacy] |M: [Ex. 6 - Personal Privacy] spradling.leann@epa.gov

-----Original Message-----
From: Spradling, LeAnn
Sent: Wednesday, February 08, 2017 3:51 PM
To: Hessling, Michael <Hessling.Michael@epa.gov>; drupal_oei_support <drupal_oei_support@epa.gov>
Cc: WCSD Webteam <WCSD_Webteam@epa.gov>
Subject: RE: Please move *.conf files to snapshot

Ok I think I see what you mean now.
I will look into it further. They should be being read - however we may be seeing conflicts between the redirects we put into the .htaccess file, the rewrite rule which rewrites the directories to _.html, and the *conf files.

Thanks,
LeAnn Spradling | Director of Consulting – Service Delivery Manager CAP/RAP EPA CAM | CGI Federal

Ex. 6 - Personal Privacy

T: [Ex. 6 - Personal Privacy] |M: [Ex. 6 - Personal Privacy] | spradling.leann@epa.gov

-----Original Message-----
From: Hessling, Michael
Sent: Wednesday, February 08, 2017 3:39 PM
To: Spradling, LeAnn <Spradling.LeAnn@epa.gov>; drupal_oei_support <drupal_oei_support@epa.gov>
Cc: WCSD Webteam <WCSD_Webteam@epa.gov>
Subject: RE: Please move *.conf files to snapshot

It did not.

In particular, https://19january2017snapshot.epa.gov/climatechange/science/index.html did not redirect
while https://www.epa.gov/climatechange/science/index.html does.

Are these *.conf rules being read?

~Mike

-----Original Message-----
From: Spradling, LeAnn
Sent: Wednesday, February 08, 2017 3:36 PM
To: Hessling, Michael <Hessling.Michael@epa.gov>; drupal_oei_support <drupal_oei_support@epa.gov>
Cc: WCSD Webteam <WCSD_Webteam@epa.gov>
Subject: RE: Please move *.conf files to snapshot

This is done, did it address the problem?

Thanks,
LeAnn Spradling | Director of Consulting – Service Delivery Manager CAP/RAP EPA CAM | CGI Federal

Ex. 6 - Personal Privacy

T: [Ex. 6 - Personal Privacy] |M: [Ex. 6 - Personal Privacy] spradling.leann@epa.gov

-----Original Message-----
From: Hessling, Michael
Sent: Wednesday, February 08, 2017 12:42 PM
To: drupal_oei_support <drupal_oei_support@epa.gov>
Cc: WCSD Webteam <WCSD_Webteam@epa.gov>
Subject: Please move *.conf files to snapshot

Hi.

Judy is uncovering broken links on snapshot.epa.gov that aren't covered by the redirects LeAnn just
added.

Like this page, https://snapshot.epa.gov/ghgreporting/forms/contact-us-about-ghg-reporting_.html -> all of
the bulleted URLs are broken, as is the EPA's Media Contacts link. That's because they're pointing to
various URLs, like /climatechange/, and redirects that are handled in the *.conf files on www.

E.g.,

https://snapshot.epa.gov/climatechange/science/index.html is 404
https://www.epa.gov/climatechange/science/index.html redirects to https://www.epa.gov/climate-change-science because of the *.conf redirect.

https://snapshot.epa.gov/climatechange/policy/index.html is 404
https://www.epa.gov/climatechange/policy/index.html is also 404 ... So whatever.

https://snapshot.epa.gov/climatechange/ghgemissions/index.html is 404
https://www.epa.gov/climatechange/ghgemissions/index.html is also 404. Hmm.

https://snapshot.epa.gov/climatechange/effects/index.html is 404. So is
https://www.epa.gov/climatechange/effects/index.html

https://snapshot.epa.gov/climatechange/wycd/index.html is 404
https://www.epa.gov/climatechange/wycd/index.html redirects to https://www.epa.gov/climatechange/what-you-can-do-about-climate-change because of the *.conf redirect.

So Judy, we can take care of some of these broken snapshot links, but others are just plumb broken on www.

~Mike